

We have as input as a synthetically generated colored image of species of foraminifera fossils and an output as 3D model. For data pre-processing, we extract the normal image, depth image and silhouette from the input 2d image of the foraminifera. The normal image is masked with the silhouette. Normal image(512x512) and depth image(512x512) are concatenated and passed to the network. Output (256x256x256) is a voxelized model and this will be our ground truth for the network.

Our network structure follows the Marrnet model which consists of encoder decoder model for 3D model generation. The enoder network with 5 convolutional 2D layers and two fully connected layers. It generates a 200 dimensional latent space vector which acts as a input to the decoder network. The decoder network consists of 5 convolutional layers that generate a 3d voxel output of size 256x256x256.